

CHAPTER II

METHODOLOGY

2.1 Part 1: Material and methods

2.1.1 Study population:

Patients who were newly admitted to the neurosurgical unit or had mental status alteration during admission were recruited. To assess the properties of the score over all types of patient, all four categories would be studied; awake, drowsy, stuporous, or comatose using the definitions previously described(25). All categories would comprise proportionate number of patients; awake 40.4%, drowsy 44.1%, stupor 8.8% and coma 6.7% according to McNarry(12) and randomly selected by the screening physician who was not involved in the score assessment during the time period.

2.1.1.1 Inclusion criteria:

2.1.1.1.1 Age over 15 years old.

2.1.1.1.2 Thai language understandable.

2.1.1.1.3 Admit to the neurosurgical ward or neurosurgical intensive care unit, Prasat Neurological Institute, the tertiary neurological health care center.

2.1.1.2 Exclusion criteria:

2.1.1.2.1 Patients affected by sedative or neuromuscular blocking agents.

2.1.1.2.2 Patients with the aftereffect from anesthesia within 24 hours.

2.1.1.3 Sample size:

Due to several main outcomes, the sample size was calculated based on 3 objectives using nQuery Advisor software version 6.

2.1.1.3.1 *The inter-observer reliability.* To evaluate the reliability between different observers, the calculation of the sample size was done using the reference from previous study(7). Assuming that the proportion

of successes was 0.5, a large sample 0.05 level two-sided test of the null hypothesis that kappa was 0.5 would have 80% power to detect an alternative hypothesis of kappa of 0.820 when the sample size was 58. With the expected 10% drop out and equal number of patient in every single level of consciousness group basis, there would be a sample of sixty four patients enrolled in this study; 26 awake, 28 drowsy, 6 stupor and 4 coma.

2.1.1.3.2 Wilcoxon's signed rank test: To test the difference in scores between two raters, i.e. expert and inexperienced raters, Wilcoxon's signed rank test was employed. However this approach was not provided by this software, so the author would use paired t-test instead. Due to no preceding study available, the calculation was done using the clinical expert concept of the FOUR score and the GCS. A sample size of 34 would have 80% power to detect a difference in means of 2, assuming a standard deviation of difference of 4, using a paired t-test with a 0.05 two-sided significance level. Twelve patients each for the assessment between expert clinicians and the other 3 assessors' classifications were accomplished; 15 awake, 16 drowsy, 3 stupor and 2 coma.

2.1.1.3.3 Spearman's rank correlation coefficients: Rank correlation was calculated to assess the correlation between the FOUR score and the GCS which measured by the expert clinicians. According to Wijdicks study(7), Spearman's correlation coefficient (r) between both scores was 0.92. However, in this study r of 0.72 was used to ensure a large enough sample size. A 0.05 two-sided Fisher's z test of the null hypothesis that the Spearman correlation coefficient (r) = 0.5, would have 80% power to detect a r of 0.72 when the sample size was 65.

2.1.2 Methods:

The FOUR and GCS score would be assessed by four types of raters, each with two personnel i.e., expert clinicians (EC1, EC2), novice clinicians (NC1, NC2), experienced registrar nurses (EN1, EN2) and apprenticed registrar nurses

(AN1, AN2). To protect patients from over-assessment, only 2 raters would independently examine and assign both FOUR and GCS score to each patient on the day of admission or the day of mental status changes. Even though expert clinicians (EC1, EC2) were found to have good to excellent agreement for both score evaluation with weighted kappa of 0.82 and 95% confidence intervals of 0.77-0.88 in the FOUR score and 0.76-0.87 in the GCS consecutively(7), their agreement would also be tested in this study. Thus, to evaluate inter-observer reliability, there would be 4 pairs of raters i.e., (1) EC1 vs. EC2; (2) NC1 vs. NC2; (3) EN1 vs. EN2; and (4) AN1 vs. AN2.

In case of good inter-observer agreement in each type of observer, one observer from each type would be randomly chose to test difference in total FOUR and GCS score between expert and inexperienced observer by selected closed envelope designated the sort of raters for each different patient i.e., (1) EC vs. NC; (2) EC vs. EN; and (3) EC vs. AN.

To reduce bias, each combination of rater would perform the examinations on the same number of patients. Intra-rater or test-retest reliability would be unable to evaluate due to impossibility of the same rater to score the same patient in such an abrupt time period without remembering the previous scores.

Prior to examining the patients, all raters would be re-educated in the FOUR and GCS scoring process using 4 lived assembly demonstrations of each consciousness level. Also provided was a handout with the figure and written tuitions describing both scores, specifying the stimuli to be used and the order in which observations would be taken. The assessment would be done within 30 minutes after the screening clinician rate the patient's consciousness level. During the patient evaluation each rater would follow the instructions and complete the scoring sheet.

To evaluate feasibility, both GCS and FOUR score would be assessed using a five-point scale (1 = strongly disagree, 5 = strongly agree) on 7 questions related to their use: (1) the score is easy to learn; (2) clinically relevant; (3) easy to use; (4) obtained in an appropriate time; (5) a potential alternative for

consciousness assessment; (6) undisturbed the patients; and (7) will be used personally if it becomes accepted in general.

2.1.3 Operative definition:

2.1.3.1 Level of consciousness(25)

2.1.3.1.1 Awake was classified if the patients opened their eyes without prompting, looked about and conversed (if they are not aphasic).

2.1.3.1.2 Drowsy was classified if the patients had their eyes closed but opened after conversational voice addressed or gently shaken, responded with sensible speech, usually became sleepy within seconds after the conversation and had natural movements of the limbs on the side without hemiparesis.

2.1.3.1.3 Stupor was classified if the patients maintained sleepiness for several minutes, closed their eyes, presented momentary or no arousal after shaking and yelling, mumbled, responded with single word or no verbal reply and consistently push aside the examiner's prodding hand with their good arms.

2.1.3.1.4 Coma was classified if the patients had no speech, still closed their eyes even vigorous shaken and made no attempt to push away the examiner's hand.

2.1.3.2 The Glasgow Coma Scale was referred to Teasdale and Jennett's article in Lancet 1994 as followed (Figure 2). The impossible evaluation of every single component of eye opening, verbal response, or motor response such as severe eyelids swelling, endotracheal intubation, tracheostomy, upper limb amputee, psychiatric disorders, developmental delay, etc. would be scored 1 each. The highest number of separable subscore would be recorded in case of hesitated determining.

Figure 2: The Glasgow Coma Scale.

Glasgow Coma Scale	
Eye opening	
4	= spontaneous
3	= to speech
2	= to pain
1	= no response
Best motor response	
6	= obeys
5	= localizes
4	= withdraws
3	= abnormal flexion
2	= extends
1	= no response
Best verbal response	
5	= oriented
4	= confused conversation
3	= inappropriate words
2	= incomprehensible sounds
1	= no response

2.1.3.3 The Full Outline of UnResponsiveness score was referred to Wijdicks and colleague's article in *Annals of Neurology* 2005 as followed (Figure 3). The impossible evaluation of every single component of eye response, motor response, brainstem reflexes, or respiration such as severe eyelids swelling, upper limb amputee, psychiatric disorders, developmental delay, etc. would be scored zero each. The highest number of separable subscore would be recorded in case of hesitated determining. Grading of respiration would be done preferably with arterial

carbondioxide partial pressure (PaCO₂) within normal limits and no ventilator adjustment while the patient was examined.

Figure 3: The Full Outline of UnResponsiveness score.

FOUR Score

Eye response

- 4 = eyelids open or opened, tracking, or blinking to command
- 3 = eyelids open but not tracking
- 2 = eyelids closed but open to loud voice
- 1 = eyelids closed but open to pain
- 0 = eyelids remain closed with pain

Motor response

- 4 = thumbs-up, fist, or peace sign
- 3 = localizing to pain
- 2 = flexion response to pain
- 1 = extension response to pain
- 0 = no response to pain or generalized myoclonus status

Brainstem reflexes

- 4 = papillary and corneal reflexes present
- 3 = one pupil wide and fixed
- 2 = papillary or corneal reflexes absent
- 1 = papillary and corneal reflexes absent
- 0 = absent papillary, corneal, and cough reflex

Respiration

- 4 = not intubated, regular breathing pattern
- 3 = not intubated, Cheyne-Stokes breathing pattern
- 2 = not intubated, irregular breathing
- 1 = breathes above ventilator rate
- 0 = breathes at ventilator rate or apnea

FOUR = Full Outline of UnResponsiveness

2.1.3.4 Category of raters

2.1.3.4.1 Expert clinicians were any physicians who had involved in neurological professional practice over or equal to 5 years.

2.1.3.4.2 Novice clinicians were any physicians who had practiced in neurological professional less than 5 years.

2.1.3.4.3 Experienced registrar nurses were any graduated nurses held certificates in neuroscience nursing and had practiced in neurological nursing at least for 2 years.

2.1.3.4.4 Apprenticed registrar nurses were any graduated nurses who had not previously worked in neurosurgical or neurological ward or newly graduated within 6 months.

2.1.4 Outcome variables:

2.1.4.1 Demographic and baseline variables: age, gender, diagnosis and level of consciousness.

2.1.4.2 The total score and subscore of the FOUR score and the GCS.

2.1.4.3 The frequency of total FOUR score and total GCS score from expert clinicians.

2.1.4.4 Total score related to practical use of the FOUR score and the GCS.

2.1.5 Statistical data analyses:

2.1.5.1 Reliability

Using data from all types of patients, inter-observer reliability of the FOUR and GCS score would be determined for each pair of rater and all pairs combined. Weighted kappa would be computed for subscale score of the FOUR score and the GCS. Similarly for each pair of rater, intraclass correlation coefficient (ICC) using a 2-way random effect model would be calculated for total score of the FOUR score and the

GCS whereas a 1-way random effect model would be used for all pairs of raters combined (Table 1).

Table 1: Inter-observer reliability using weighted kappa and intraclass correlation.

Pair of rater	n	FOUR score				GCS score				
		Eye	Motor	Brainstem	Respiration	Total	Eye	Motor	Verbal	Total
EC1, EC2										
NC1, NC2										
EN1, EN2										
AN1, AN2										
Total										

EC = Expert clinician, NC = Novice clinician, EN = Experienced nurse, AN = Apprenticed nurse.

Since the FOUR score and the GCS from expert clinician (EC) were considered to be correct, the differences in total FOUR and total GCS score between expert and inexperienced raters i.e., EC vs. NC, EC vs. EN, and EC vs. AN would be analyzed using Wilcoxon's signed rank test (Table 2).

Table 2: The difference of the FOUR score and the GCS using Wilcoxon's signed rank test.

Pair of rater	Category of rater	n	Total FOUR score			Total GCS score		
			Mean	SD	p-value	Mean	SD	p-value
EC, NC	EC (1 or 2)							
	NC (1 or 2)							
EC, EN	EC (1 or 2)							
	EN (1 or 2)							
EC, AN	EC (1 or 2)							
	AN (1 or 2)							

EC = Expert clinician, NC = Novice clinician, EN = Experienced nurse, AN = Apprenticed nurse.

Cronbach's alpha would be determined for the FOUR score and the GCS to assess internal consistency using score from the first and second rater of expert clinicians (Table 3).

Table 3: Internal consistency of the FOUR score and the GCS using Cronbach's alpha.

Raters	n	FOUR score	GCS score
First expert clinician			
Second expert clinician			

2.1.5.2 Validity

2.1.5.2.1 Content validity was as previously mentioned in Background and rationale part.

2.1.5.2.2 Construct validity of the FOUR score would be evaluated using Spearman's rank correlation coefficient between the total FOUR and total GCS score from the first and second rater expert clinicians (Table 4).

Table 4: Spearman's rank correlation between the total FOUR and the total GCS score.

Raters	n	Spearman's rank correlation	p-value
First expert clinician			
Second expert clinician			

2.1.5.2.3 Criterion validity of the FOUR score should be assessed by sensitivity and specificity analysis. Nonetheless, this study did not mention the "Gold standard" for the mortality prediction. In order to approximate mortality from the FOUR score, the cut points were estimated from statistical value as

demonstrated in Table 5. Afterward, the weighted kappa between the total FOUR score at the estimated cut points and the total GCS at two different cut points; ≤ 6 , >6 and <12 , ≥ 12 were evaluated (Table 6).

Table 5: The cut points of the FOUR score estimated by the rank of GCS.

FOUR score	GCS score
Mean	
SD	
Min-max	

Table 6: The reliability of the FOUR score cut points compared to the GCS cut points using weighted kappa.

FOUR score	n	GCS score		
		3-6	7-11	12-15
0 – first cut point				
first – second cut point				
Second cut point – 16				

2.1.5.3 Practical feasibility

Feasibility of the FOUR and GCS score based on a subscore and total score from 7 questions (each with a 5-point Likert scale) would be compared using Wilcoxon's signed-rank test as shown in Table 7.

Table 7: Feasibility of the FOUR score compare to the GCS using Wilcoxon's signed rank test.

Questions	FOUR score		GCS score		p-value
	Mean (SD)	Min-max	Mean (SD)	Min-max	
1. The score is easy to learn					
2. Clinically relevant					
3. Easy to use					
4. Obtained in an appropriate time					
5. A potential alternative for consciousness assessment					
6. Undisturbed the patients					
7. Will be used personally if it becomes accepted in general					

2.2 Part 2: Results

There were 100 patients or 200 patients' records available for final analysis. The mean age of the study sample was 53.7 ± 15.9 years with a range of 15-89 years. Fifty-two percent were female (52 patients) and 48% were male (48 patients). The distribution of the patients' diagnoses were brain tumour (48%), intracranial hemorrhage (22%), intracranial aneurysm (11%), spondylolithesis (6%), hydrocephalus (3%), spinal cord tumour (3%), intracranial infection (2%), pneumocephalus (2%), trigeminal neuralgia (1%), herniated nucleus pulposus (1%) and skull defect (1%). The patients were categorized into 4 stages of consciousness as followed; 41% awake, 44% drowsy, 9% stupor, and 6% coma. Twenty percent of the patients were intubated and mechanically ventilated.

2.2.1 Reliability analysis:

The overall reliabilities of each pair of rater were excellent for both the total FOUR score and the GCS, intraclass correlations (ICC) were in 0.90-0.99 range as demonstrated in Table 8 and Table 9. The Apprenticed nurses were the most consistent in their assessment (ICC = 0.99, 0.97) in the total FOUR score and the GCS

respectively. The lowest degree of agreement was among the novice clinician group but the values were still excellent in both total scores (ICC = 0.90). Within group reliability of the FOUR score showed poor level of agreement for brainstem subscale in expert clinician, novice clinician and experienced nurse group while the reliability for eye subscale in the GCS was lowest in expert clinicians.

Table 8: Inter-observer reliability of the FOUR score demonstrated in Quadratic weighted kappa for each subscale and intraclass correlation for the total score.

Pair of rater	n	FOUR score				
		Eye	Motor	Brainstem	Respiration	Total
		(0-4)	(0-4)	(0-4)	(0-4)	(0-16)
EC1, EC2	16	0.40	1.00	0	0.88	0.93
		(0.07-0.73)	(1.00-1.00)	(0-0)	(0.64-1.00)	(0.80-0.97)
NC1, NC2	16	0.92	0.54	0	1.00	0.90
		(0.84-1.00)	(0.30-0.77)	(0-0)	(1.00-1.00)	(0.72-0.97)
EN1, EN2	16	0.88	0.55	0	0.62	0.94
		(0.76-1.00)	(0.27-0.83)	(0-0)	(0.15-1.00)	(0.84-0.98)
AN1, AN2	16	0.86	0.81	0.98	0.99	0.99
		(0.70-1.00)	(0.65-0.97)	(0.93-1.00)	(0.96-1.00)	(0.97-1.00)

FOUR score = Full outline of unresponsiveness score, GCS = Glasgow coma scale score, EC = Expert clinician, NC = Novice clinician, EN = Experienced nurse, AN = Apprenticed nurse.

Table 9: Inter-observer reliability of the GCS score demonstrated in weighted kappa for each subscale and intraclass correlation for the total score.

Pair of rater	n	GCS score			
		Eye	Motor	Verbal	Total
		(1-4)	(1-6)	(1-5)	(1-15)
EC1, EC2	16	-0.07 (-0.18-0.05)	1.00 (1.00-1.00)	0.98 (0.94-1.00)	0.97 (0.91-0.99)
NC1, NC2	16	0.59 (0.35-0.83)	0.11 (-0.20-0.42)	0.95 (0.90-1.00)	0.90 (0.71-0.96)
EN1, EN2	16	0.69 (0.43-0.94)	0.44 (0.18-0.69)	0.85 (0.67-1.00)	0.95 (0.87-0.98)
AN1, AN2	16	0.73 (0.43-1.00)	0.84 (0.64-1.00)	0.93 (0.86-1.00)	0.97 (0.91-0.99)

FOUR score = Full outline of unresponsiveness score, GCS = Glasgow coma scale score, EC = Expert clinician, NC = Novice clinician, EN = Experienced nurse, AN = Apprenticed nurse.

There were no statistically significant differences in the total FOUR score and the GCS between expert clinicians and the others, i.e. NC, EN, AN, as indicated by Wilcoxon's signed rank test (Table 10).

Table 10: The difference of the total FOUR score and the total GCS between group evaluators using Wilcoxon's signed rank test demonstrated in mean, standard deviation (SD) and p-value.

Pair of rater	Category of rater	n	Total FOUR score			Total GCS score		
			Mean	SD	p-value	Mean	SD	p-value
EC, NC	EC (1 or 2)	12	12.58	5.37		11.33	4.66	
	NC (1 or 2)	12	11.08	5.04	0.59	11.08	5.04	0.41

EC, EN	EC (1 or 2)	12	14.42	1.88		13.25	2.30	
	EN (1 or 2)	12	15.08	1.56	0.046	13.42	2.23	0.59
EC, AN	EC (1 or 2)	12	15.17	2.13		13.67	2.27	
	AN (1 or 2)	12	14.83	2.13	0.16	13.67	2.42	1.00

FOUR score = Full outline of unresponsiveness score, GCS = Glasgow coma scale score, EC = Expert clinician, NC = Novice clinician, EN = Experienced nurse, AN = Apprenticed nurse.

The Cronbach's alpha value reflecting internal consistency was good (>0.80) for both the FOUR and the GCS score evaluated by the expert clinicians (Table 11).

Table 11: Internal consistency of the FOUR score and the GCS evaluated by expert clinicians demonstrated in Cronbach's alpha value.

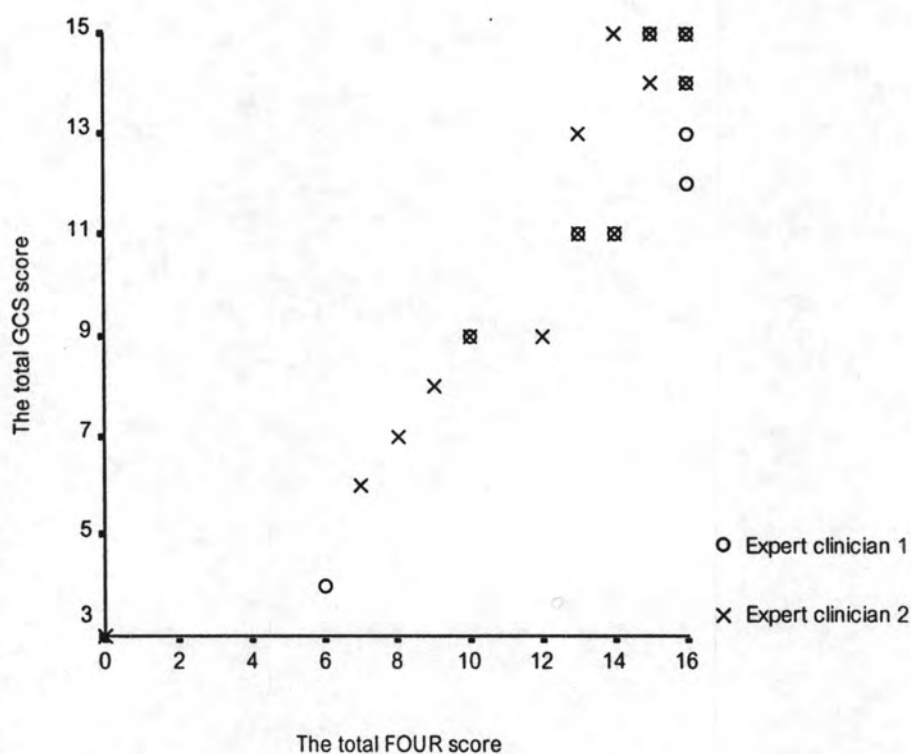
Raters	n	FOUR score	GCS score
First expert clinician	38	0.83	0.85
Second expert clinician	30	0.83	0.83
	68	0.82	0.85

FOUR score = Full outline of unresponsiveness score, GCS = Glasgow coma scale score

2.2.2 Validity analysis:

Emphasize on the construct validity of the total FOUR score, the author examined scatter plot between the total FOUR score and the total GCS score with Spearman's rank correlation coefficient. The correlation was good ($r = 0.78$) with p-value <0.01 (Figure 4).

Figure 4: Scatter plot of the total FOUR score and the total GCS score in expert clinician group.



FOUR score = Full outline of unresponsiveness score, GCS = Glasgow coma scale score

The criterion validity of the total FOUR score was preliminary estimated by the statistical value cut points according to the total score of the GCS at 3-6, 7-11 and 12-15. Table 12 displayed mean, SD, minimum and maximum of the total FOUR score for each level of the total GCS score. The best reliability of the total FOUR score cut points compared with the total GCS indicated by weighted kappa (K_w) was 0.92 with 95% confidence interval 0.83-1.00 at 0-7, 8-14 and 15-16 range. The frequency was shown in Table 13. Box and Whiskers plot of the range of the total GCS encompassed by different cut points of the total FOUR score (0-7, 8-14, 15-16) was demonstrated in Figure 5.

Table 12: The cut points of the FOUR score estimated by the rank of GCS.

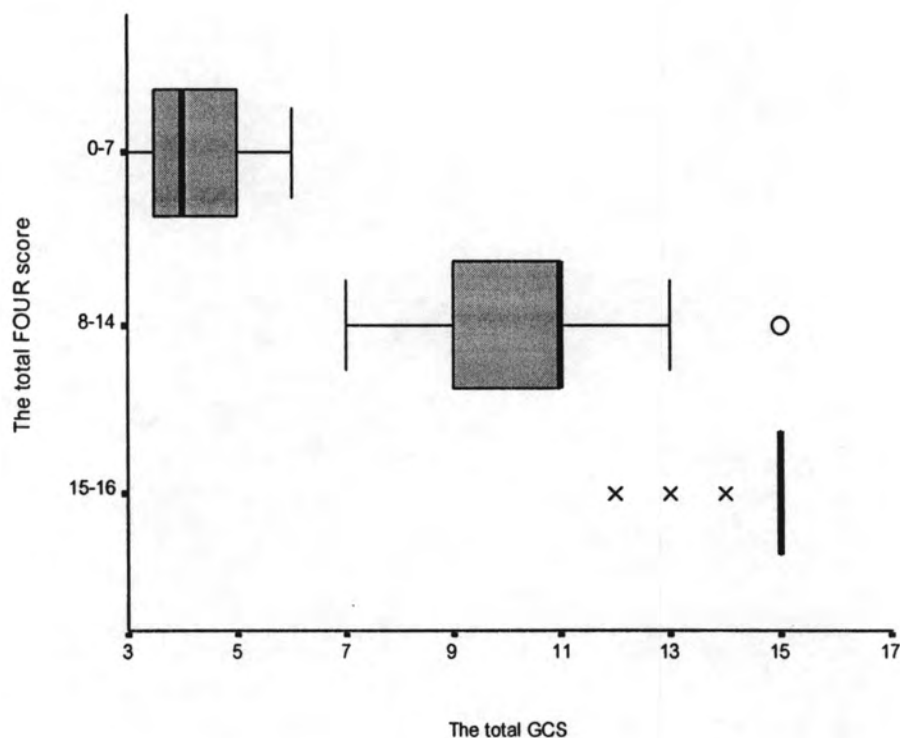
FOUR score	GCS score		
	3-6	7-11	12-15
Mean	4.33	11.64	15.76
Median	6	12	16
SD	3.79	2.06	0.67
Min-max	0-7	8-14	13-16

FOUR score = Full outline of unresponsiveness score, GCS = Glasgow coma scale score, Min = minimum, Max = maximum

Table 13: The frequency of the patients categorized by the total FOUR score cut points compared to the GCS cut points.

FOUR score	GCS score		
	3-6	7-11	12-15
	(High risk)	(Intermediated risk)	(Low risk)
0 – 7 (High risk)	3	0	0
8 – 14 (Intermediated risk)	0	11	3
15 – 16 (Low risk)	0	0	51

Figure 5: Box and Whiskers plot of the total GCS for different categories of the total FOUR score.



2.2.3 Practical feasibility analysis:

The practicability analysis of the FOUR score and the GCS was accomplished from 200 case recorded forms by 8 examiners at different episodes. Amongst 7 questions, the lowermost scores were the question related to the patients' undisturbed phenomenon, 3.81 ± 0.98 in FOUR score and 4.13 ± 0.97 in the GCS. The overall optimistic scores of the GCS became outstanding over the FOUR score significantly (Table 14). The total practical feasibility score means were 4.29 in FOUR score and 4.72 in the GCS even as the variances were 0.47 and 0.07 respectively.

Table 14: Practical feasibility score of the FOUR score compared to the GCS using 5 point Likert's scale (1 = strongly disagree, 5 =strongly agree).

Questions	FOUR score		GCS score		p-value
	Mean (SD)	Min- max	Mean (SD)	Min- max	
1. The score is easy to learn	4.39 (0.63)	2-5	4.86 (0.39)	2-5	<0.01
2. Clinically relevant	4.42 (0.68)	2-5	4.82 (0.39)	4-5	<0.01
3. Easy to use	4.31 (0.68)	2-5	4.83 (0.42)	2-5	<0.01
4. Obtained in an appropriate time	4.38 (0.63)	2-5	4.79 (0.46)	3-5	<0.01
5. A potential alternative for consciousness assessment	4.28 (0.73)	1-5	4.73 (0.67)	1-5	<0.01
6. Undisturbed the patients	3.81 (0.98)	1-5	4.13 (0.97)	1-5	<0.01
7. Will be used personally if it becomes accepted in general	4.41 (0.67)	2-5	4.88 (0.34)	3-5	<0.01

FOUR score = Full outline of unresponsiveness score, GCS = Glasgow coma scale score